

**EDITOR'S SUMMARY**

As of January 2011, the National Science Foundation (NSF) has required a data management plan (DMP) as part of all grant proposals. Georgia Tech's librarians and information technologists presented DMP development guidelines to support the requirement and later examined 181 plans submitted in the first eight months of 2011. The review revealed only 22 percent referred to Georgia Tech's institutional repository, SMARTech. The schools submitting the greatest percentage of proposals were not most likely to reference SMARTech, nor was there consistency in referencing the university repository within schools. The language used to describe SMARTech was highly variable, drawn from various sources and widely shared. Public online collaboration tools were also referenced in proposals. The analysis indicates inconsistent understanding of the university's data repository services, the need for consistent descriptive language and the need to target specific schools for outreach, all in partnership with researchers.

**KEYWORDS**

digital repositories  
colleges and universities  
scientific and technical information  
research data sets  
proposals  
grants

## NSF DMP Content Analysis: What Are Researchers Saying?

by Susan Wells Parham and Chris Doty

The National Science Foundation (NSF) implemented its requirement that all grant proposals include a data management plan (DMP) in January 2011. Like our colleagues at research institutions across the United States, librarians and technologists at Georgia Tech developed services to support this mandate, including guidelines and workshops for developing a DMP. Toward the end of the requirement's first year, we assessed the impact of our consultation and outreach services by reviewing the content of submitted data management plans.

In cooperation with the GT Office of Sponsored Programs, we examined NSF DMPs submitted by Georgia Tech researchers during the first eight months of the mandate (through September 6, 2011). Of the 335 submitted proposals, we reviewed the content of 181 plans. We excluded those proposals that were grant supplements or transfers. Using plagiarism software, we searched DMP content for information related to repository services, inter- and intradepartmental sharing

of DMPs and the prevalence of cloud-based tools. This brief article outlines our findings and their influence on strategic planning for a range of research data curation services, including data repository services and related data stewardship initiatives.

### Repository Services

Of the 181 NSF DMPs that were analyzed, 39 (22%) identified Georgia Tech's institutional repository, SMARTech. The percentage of plans which included SMARTech varied widely by school. For example, five researchers from mechanical engineering (ME) referenced SMARTech services in their plans. This number is approximately 14% of the total DMPs submitted from the school. On the other hand, eight researchers from the much smaller school of aerospace engineering mentioned SMARTech (approximately 62% of total submitted plans).

The breakdown of SMARTech DMP language by school allowed us to evaluate the effectiveness of our campus outreach and to prioritize outreach in the future. In only one school did researchers mention SMARTech in over 50% of the DMPs. Five schools did not mention repository services at all. The top three schools account for almost half of all NSF proposals over the time period studied, and they identify SMARTech in only 20% of DMPs. While inconsistent outreach may not be

Susan Wells Parham is head of scholarly communication & digital curation at the Georgia Tech Library. She can be reached at [susan.parham@at.library.gatech.edu](mailto:susan.parham@at.library.gatech.edu).

Chris Doty is librarian for physics and materials science & engineering at the Georgia Tech Library. He can be reached at [chris.doty@at.library.gatech.edu](mailto:chris.doty@at.library.gatech.edu).

the primary factor driving these numbers, we can clearly target particular schools, such as ME. It's not only the largest school on campus; it also has the greatest number of NSF proposals awarded.

Researchers obtained language pertaining to SMARTech from a number of locations. Nine DMPs contained language from SMARTech's "About" webpage. Eight more contained language from SMARTech's "Mission and Collection Policy" webpage. Four DMPs contained language about SMARTech found on the library website, and one had language from the Georgia Tech Faculty Handbook. Five DMPs contained substantial language about SMARTech whose source was not easily identified; presumably it was original to the authors. Twelve of the 39 DMPs that identified SMARTech contained only the briefest of mentions, in a pair of cases only the web address was given.

Identifying where this language in the DMPs originated was important because it illustrated multiple, often inconsistent sources of information. Clearly, when we update language regarding repository services, it will need to be modified in multiple other locations as well. The review led us to change the text on the SMARTech webpage to explain more clearly its role in long-term storage. This consistency is particularly important as we implement a new digital preservation strategy.

### DMP Sharing

Sharing of DMP text among faculty members was relatively common and often occurred across departmental boundaries. One third of faculty members had large sections of text identical to at least one other researcher's DMP. Half of the instances of sharing were

among faculty members in different schools. Two thirds of shared DMP text was between just a pair of faculty members, while the other third consisted of groups of four, five and six different faculty members. Because researchers are obviously sharing a significant amount of DMP text, we need to ensure that they have consistent, up-to-date language about repository services. We'll need to distribute this text widely to counteract the widespread use of outdated boilerplate language.

### Cloud-based Tools

Dropbox was mentioned by five faculty members. Three of them proposed using Dropbox to facilitate collaboration with colleagues at Georgia Tech and abroad. One of three also proposed using Google Docs in a similar manner. In another case, Dropbox was listed as one of a handful of possible venues for sharing data (iCloud and Amazon were others). One faculty member proposed using Dropbox as long-term backup for data housed on laboratory servers. Two faculty members also proposed using Google Code as an open source repository for code, documentation, and input and output data.

During this study, we discovered both a lack of information about repository data services and the widespread sharing of text describing a onetime digital preservation model. We have a clear road ahead of us: we will target specific schools for outreach; develop consistent language about repository services for research data; and focus on the widespread dissemination of information about our new digital preservation strategy. To succeed, this effort will need to be a partnership among librarians, administrators, technologists and researchers themselves. ■